SEX DETERMINATION FROM FINGERPRINT

R. A. Tadross[1], A. M. Badawi[1], M. R. Mahfouz[1], R. L. Jantz[2], C. M. Blair[1]
[1]Center for Musculoskeletal Research, University of Tennessee, Knoxville, TN, USA
[2]Anthropology Department, University of Tennessee, Knoxville, TN, USA
e-mail: mmahfouz@utk.edu

Abstract- **In forensic anthropology, gender classification from fingerprints is an important step when identifying the gender of a suspected criminal. A dataset of 10-fingerprint images for 2200 persons of different ages and gender (1100 males and 1100 females) was analyzed. Features extracted included ridge count, ridge thickness to valley thickness ratio (RTVTR), white lines count, ridge count asymmetry, and pattern type concordance. Fuzzy C-Means (FCM), Linear Discriminant Analysis (LDA), and Neural Network (NN) were used for the classification using the most dominant features. We obtained results of 79.2%, 85.44%, and 87.34% using FCM, LDA, and NN, respectively. Results of this analysis make this method a prime candidate to utilize in forensic anthropology for gender classification in order to minimize the suspects search list by getting a likelihood value for the criminal gender.**
***Keywords* - Gender classification, fingerprint, white lines, forensic anthropology.**

I. INTRODUCTION

Fingerprint identification and classification has been extensively researched in the literature [1]; however, very few researchers have studied the fingerprint gender classification problem [2-13]. Acree examined the ridge density, defined as the number of ridges in a certain space; showing that females have a higher ridge density compared to males [2]. Kralik demonstrated that the males have higher ridge breadth, defined as the distance between the centers of two adjacent valleys, than females [3]. Two studies showed that the males have higher ridge count than females [4-5]. Also, it has been reported that males and females have higher rightward directional asymmetry in the ridge count [5-8], with the asymmetry being higher in males than females [8], and higher incidence of leftward asymmetry in females [5]. Female's fingerprints are significantly of lower quality than male fingerprints [9]; however, the appearance of white lines and scars in fingerprint images is very common in females [10]. In this paper we studied the different debates in the literature for the few articles that exist concerning the significance of ridge count, pattern type concordance, ridge count asymmetry, ridge thickness to valley thickness ratio (RTVTR), and white lines count features on the classification performance. We analyzed different features that can be significant in gender classification and different classifiers performances.

II. MATERIALS AND METHODOLOGY

In our gender classification analysis from fingerprints, we acquired the data before extracting the whole features for every finger. We then averaged the features for the person's 10 fingers, and classified the gender of each person based on different combinations of these features. The overall features include ridge count, RTVTR, fingerprint pattern type, white lines count, pattern type concordance between the corresponding left-right fingerprints, and ridge count asymmetry between the left-right corresponding fingerprints. Statistical analysis was performed for pattern types, ridge count, and ridge counts along pattern types.

A. Dataset

A dataset of 10-fingerprints for 2200 persons of different ages and gender (1100 males, and 1100 females) were scanned from their ink prints as shown in Figs. 1 and 2, and were analyzed for the ridge count, and pattern type features. The RTVTR, and white lines count features were analyzed for 452 persons (254 males, and 198 females).

B. Features Extraction

B.1. Ridge and Valley Thicknesses

The average ratio between the ridge thickness and the valley thickness for each of the subject's fingerprints was calculated automatically for every subject. The fingerprint image was divided into 30x30 non-overlapping blocks. The local ridge orientation within each block was calculated [16] as shown in Fig. 3. The projection profile of the valleys and ridges along a line perpendicular to the local ridge orientation in each block was calculated, and the projection profile was binarized using 1D optimal thresholding [17]. The resultant binary profile represents the ridges and valleys in this block, the high binary value represents the valleys and the low binary value represents the ridges.

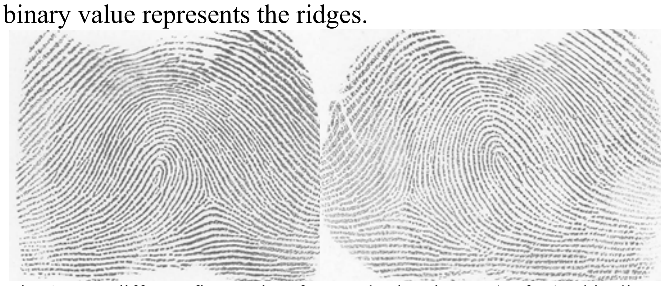
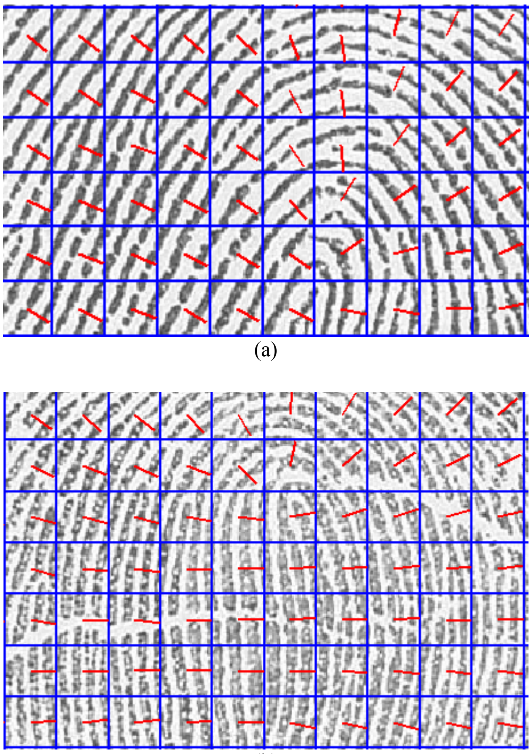


Fig. 1. Two different fingerprints for a male showing no (or few) white lines and small RTVTR.

Fig. 2. Two different fingerprints for a female showing large count of white lines and large RTVTR.

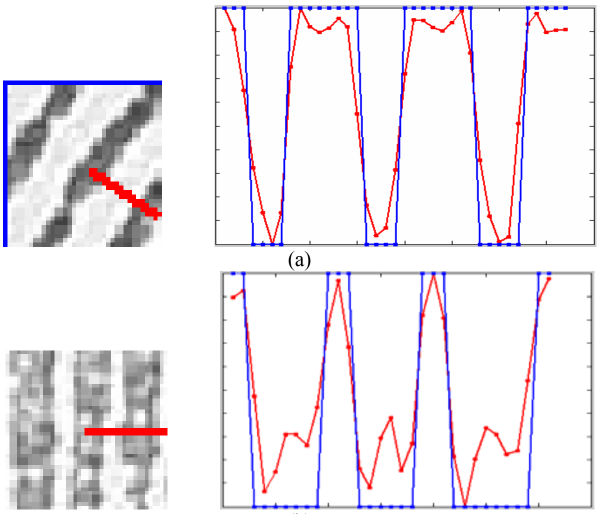


(a)

(b)

Fig. 3. (a) a male fingerprint, and (b) a female fingerprints

Fig. 4 illustrates two blocks of female and male fingerprints, and their projection and binary profiles. The average RTVTR was calculated for each block. The uniformity of ridges and valleys within the blocks varies, for blocks having non uniform ridges and valleys due to the low quality of the fingerprint image in this region, the ridge orientation estimation is usually incorrectly estimated, and thus the RTVTR calculated for this block is incorrect, so only the blocks having the best quality should contribute to the average RTVTR calculated for this fingerprint.



(a)

(b)

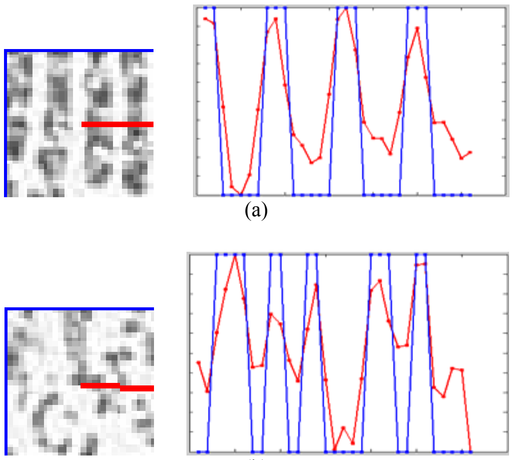
Fig. 4. Block from (a) a male's fingerprint with RTVTR of 0.54, and (b) a female's fingerprint with RTVTR of 2.33.

For each block, a quality index was calculated as the average difference between the values of successive singular points (minima and maxima) of the projection profile. Blocks of good quality have higher quality index than those of bad quality. Fig. 5 shows a good quality block, having quality index of 0.244, and correctly estimated RTVTR of 1.67, as well as a bad quality block, having quality index of 0.061, and incorrectly estimated RTVTR of 1.06. The blocks were arranged in a descending order based on their quality index, and the RTVTR of the best 10 were averaged and taken as the average RTVTR for this fingerprint.

B.2. White Lines Count, Ridge Count, Pattern Type, Pattern Type Concordance, and Ridge Count Asymmetry

The white lines count and ridge count were extracted manually for each fingerprint, an average white lines count as well as the ridge count was calculated for each subject. Pattern type was extracted manually for each fingerprint, and the pattern type concordance was calculated for the fingerprints of each left-right corresponding fingerprint pair for the subject, such that the concordance value is 1 if the corresponding fingerprints have the same pattern type, and is 0 otherwise, then the sum of the 5 fingerprint pairs concordance values was calculated.

The ridge count asymmetry between the left-right corresponding fingerprints for a subject was calculated, the asymmetry is 1 for a left-right corresponding fingerprint pair if the ridge count of the left fingerprint is greater than the right one, is –1 if it is smaller, and is 0 if both ridge counts are equal. The sum of the asymmetry values of the 5 fingerprint pairs of the subject was calculated.



(a)

(b)

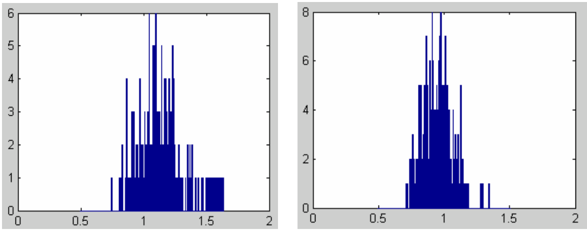
Fig. 5. (a) Good, and (b) bad quality blocks, and their profiles.

III. Results

A. RTVTR and White Lines Count Statistics

Primarily, the female's fingerprint is characterized by a high RTVTR, while the male's fingerprint is characterized by low RTVTR. Fig. 6 shows histograms of the RTVTR of the females, with μ=1.1042, σ=0.1481, and the males, with

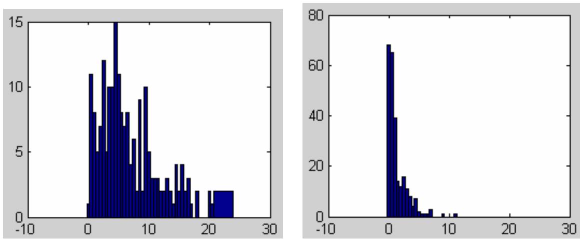
μ=0.9505, σ=0.1145, with t-value=12.44, and ρ-value=1.552e-30. The following histograms have a horizontal axis of the feature values and a vertical axis of the number of cases sharing the same feature value.



(a) (b)

Fig. 6. Histograms of the RTVTR for (a) females and (b) males.

A high count of white lines is usually associated with a female's fingerprint, with the exception of a small percentage having few or no white lines. The male's fingerprint is characterized by having no or few number of white lines, with the exception of small percentage having high count of white lines. Fig. 7 shows histograms of the white lines count for the females, with μ= 6.8859, σ=4.9493, and the males, with μ=1.326, σ=1.647, with t-value=16.76, and ρ-value=2.144e-48.

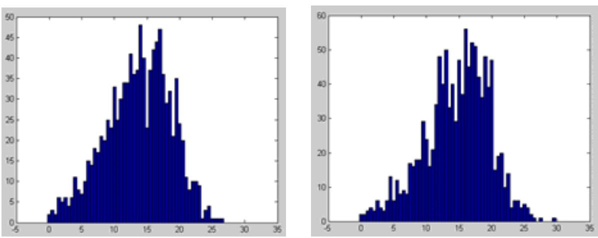


(a) (b)

Fig. 7. Histograms of white lines count for (a) females and (b) males.

B. Ridge Count, Ridge Count Asymmetry, and Pattern Type Concordance Statistics

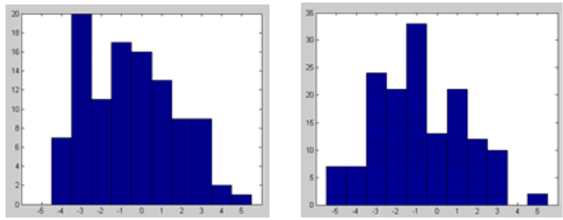
The male's ridge count is slightly higher than the female's, with high standard deviation in its distribution among both genders. Histograms of the ridge count for the females, with μ=13.6671, σ=4.9845, and the males, with μ=14.6914, σ=4.9336, with t-value =4.802, and ρ-value=1.685e-06 are shown below in Fig. 8..



(a) (b)

Fig. 8. Histograms of the ridge count for (a) the females and (b) the males.

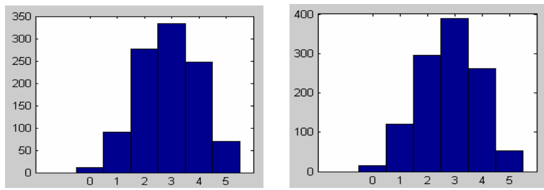
The ridge count asymmetry between left and right hand fingerprints shows high variance value for both females and males, and slight difference in the mean value between females and males, with the females having slightly higher degree of rightward asymmetry. Fig. 9 shows histograms of the asymmetry values for the females, with μ=-0.5333, σ=2.2148, and the males with μ=-0.8333, σ=2.2056, with t-value=3.155, and ρ-value=0.00163.



(a) (b)

Fig. 9. Histograms of the ridge count asymmetry for (a) the female and (b) the males..

The pattern type concordance doesn't show significant variation among male and female fingerprints. The histograms in Fig. 10 show the concordance values for the females, with t-value=2.8991, σ=1.1049, and the males with μ=2.8104, σ=1.09, with t-value =1.879, and ρ-value=0.06041.



(a) (b)

Fig. 10. Histograms of the pattern type concordance for (a) the female and (b) the males.

C. Pattern Type Statistics

The overall number of fingerprints belonging to the different 25 fingerprint types shown in Figs. 11 and 12 was analyzed; the Ulnar loop type is the most abundant type, followed by the Monocentric whorl. The probability of occurrence for each pattern type is nearly the same for males and females with slight differences, as shown in Fig. 11.

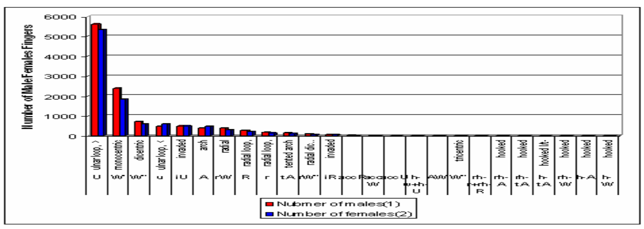


Fig. 11. Male/Female number of fingerprints per pattern type.

The ridge count varies among the pattern types. The mean ridge count for each pattern type varies slightly between males and females, as shown in Fig. 12.

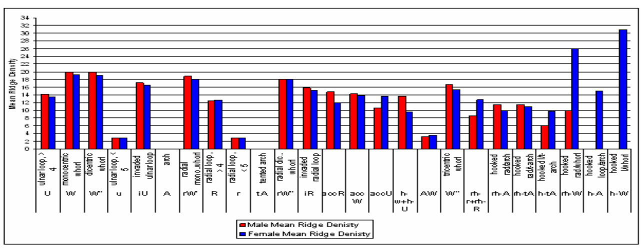


Fig. 12. Male/Female mean ridge count.

D. Classification

First we applied Fuzzy C-Means as an unsupervised clustering method for overall feature vectors, then we performed linear discriminant analysis (LDA) on the data and finally applied a Neural Network classifier on different combinations of the extracted features. We found that the most significant features are the white lines count and RTVTR. The ridge count proved to have little significance, and slightly enhance the classification with the Neural Network classifier, while it degrades the performance of the Fuzzy C-Mean classifier. Adding any of the other least significant features to the input vectors result in degradation of performance of the classifiers.

D.1. Fuzzy C-Means (FCM) Clustering

By applying the FCM algorithm on the white lines count and RTVTR only as input features, a result of 79.26% classification rate was obtained and this result is shown as a confusion matrix in Table I. Table II shows the addition of ridge count to the previous two features, with a classification of 58.4%; proving that adding ridge count to the feature vector highly decreased the classification rate of males and slightly decreased the females' classification rate.

TABLE I
CONFUSION MATRIX FOR THE FCM CLASSIFICATION BASED ON THE WHITE LINES COUNT, AND RTVTR FEATURES.

Actual\Estimated	Males	Females	Total
Males	244	10	254
Females	84	114	198
Total	328	124	452

TABLE II
CONFUSION MATRIX FOR THE FCM CLASSIFICATION BASED ON THE WHITE LINES COUNT, RTVTR, AND RIDGE COUNT FEATURES.

Actual\Estimated	Males	Females	Total
Males	148	106	254
Females	82	116	198
Total	230	222	452

D.2. Linear Discriminant Analysis (LDA)

By applying the LDA on the white lines count and RTVTR, we obtained a result of 84.81% classification for the testing set, as shown in the confusion matrix in Table III, with training error rate of 0.179. By adding the ridge count to the previous features, we attained a result of 85.44%, as shown in Table IV, with error rate of 0.171.

TABLE III
CONFUSION MATRIX FOR THE TESTING SET FOR THE LDA BASED ON WHITE LINES COUNT, AND RTVTR.

Actual\Estimated	Males	Females	Total
Males	82	3	85
Females	21	52	73
Total	103	55	158

TABLE IV
CONFUSION MATRIX FOR THE TESTING SET FOR THE LDA BASED ON WHITE LINES COUNT, RTVTR, AND RIDGE COUNT

Actual\Estimated	Males	Females	Total
Males	82	3	85
Females	20	53	73
Total	102	56	158

D.3. Neural Network Classification

A Multi Layer Back Propagation Neural Network as a non-linear classifier [13] was used for this gender classification analysis. Our data was divided into 2/3 for training and 1/3 for testing and we studied the effect of different combinations of the extracted features on the classification rate. The best number of hidden neurons for each features combination was determined empirically. The RTVTR and the white lines count features were used as the inputs for the Neural Network with 15 hidden layer neurons, learning rate of 0.2, and momentum term of 0.5. The classification rate for the training set was 87.415%**,** and 86.076% for the testing set**,** with training root mean square error (RMS) of 0.098 in 3000 epochs. Results of the training and testing confusion matrices are shown in Tables V and VI, respectively.

TABLE V
CONFUSION MATRIX FOR TRAINING SET FOR THE NEURAL NETWORK CLASSIFICATION BASED ON THE WHITE LINES COUNT, AND RTVTR FEATURES.

Actual\Estimated	Males	Females	Total
Males	153	13	166
Females	24	104	128
Total	177	117	294

TABLE VI
CONFUSION MATRIX FOR THE TESTING SET FOR THE NEURAL NETWORK CLASSIFICATION BASED ON THE WHITE LINES COUNT, AND RTVTR FEATURES.

Actual\Estimated	Males	Females	Total
Males	79	9	88
Females	13	57	70
Total	92	66	158

We found that using the RTVTR, the white lines count, and the ridge count features as inputs to the Neural Network slightly improves the performance giving 89.1156% for the training set, and 87.3418% for the testing set, with training root mean square error of 0.0793 in 43000 epochs, learning rate of 0.2, momentum term of 0.5, hidden layer of 25 neurons. The training and testing confusion matrices are shown in Tables VII, and VII.

TABLE VII
CONFUSION MATRIX FOR TRAINING SET FOR THE NEURAL NETWORK CLASSIFICATION BASED ON THE WHITE LINES COUNT, RTVTR, AND RIDGE COUNT FEATURES.

Actual\Estimated	Males	Females	Total
Males	155	11	166
Females	21	107	128
Total	176	118	294

TABLE VIII
CONFUSION MATRIX FOR THE TESTING SET FOR THE NEURAL NETWORK CLASSIFICATION BASED ON THE WHITE LINES COUNT, RTVTR, AND RIDGE COUNT FEATURES.

Actual\Estimated	Males	Females	Total
Males	77	11	88
Females	9	61	70
Total	86	72	158

Adding asymmetry, and/or concordance features was found to decrease the classification performance significantly. This is due to the fact that these features do not show significant statistical variations between males and females.

IV. CONCLUSION

Average statistics along fingers for every pattern type was calculated together with different concordance and asymmetry properties for the corresponding fingers. The variation among females and males in the membership of the fingerprints to the different pattern types, and the average ridge count for fingers belonging to each pattern type, are very small, and thus are statistically insignificant. Male and female fingerprints are characterized by an average rightward asymmetry in the ridge count, i.e. the ridge count of a finger in the right hand is most likely greater than the ridge count of its corresponding finger in the left hand, but there is no significant difference in the degree of asymmetry between males and females, and thus the asymmetry is not a good candidate for the classification process.

The pattern type concordance between left and right corresponding fingers doesn’t show significant statistical variations between females and males. We found that the most significant features are the RTVTR and the white lines count averaged over the individual’s 10 fingerprints, with the females having higher white lines count and RTVTR than the males. These two features have shown high significance in the classification process using FCM and Neural Networks classifiers. NN classifier has a higher classification rate than LDA. The average ridge count is slightly higher in males than in females, with high standard deviation among subjects of both genders. We found that adding this feature to the white lines count, and RTVTR slightly improves the performance of the classification process using Neural Network and LDA, while it degraded the performance of the FCM classifier. Highest gender classification rate was 87.3% using NNs and 85.44% using LDA. Gender classification results using these dominant features demonstrated that this method could be considered as a prime candidate for use in forensic anthropology in order to minimize the suspects search list and give a likelihood probability value of the gender of a suspect.

REFERENCES

[1] D. Maltoni, A. Jain, “Biometric Authentication”, Proceedings of ECCV 2004 International Workshop, BioAW 2004, Prague, Czech Republic, Springer, May 2004.

[2] M. Acree “Is there a gender difference in fingerprint ridge density?.”, Federal Bureau of Investigation, Washington, DC 20535-0001, USA.

[3] M. Kralik, V. Novotny” Epidermal ridge breadth: an indicator of age and sex in paleodermatoglyphics.”, Variability and Evolution, Vol. 11: 5–30, 2003.

[4] J. Hall and D. Kimura “Dermatoglyphic Asymmetry and Sexual Orientation in Men.” ,Behavioral Neuroscience, 108, 1203-1206, 1994.

[5] K. Cote, M. Christopher,and M. Lalumiere “Birth Order, Birth Interval, and Deviant Sexual Preferences Among Sex Offenders.”, Sexual Abuse: A Journal of Research and Treatment, Vol. 14, No. 1, January 2002.

[6] S. Saha, D. Loesch, D. Chant, J. Welham, O., El Saadi, L. Fañanás, B. Mowry and J. McGrath “Directional and fluctuating asymmetry in finger and a-b ridge counts in psychosis: a case-control study.” BMC Psychiatry, 3:3, 2003.

[7] J. Pascoe, “Outlooks on Homosexuality”, Ph.D., Chittenden 7-7 (Burlington & South Burlington, Vermont) "http://www.boskydell.com/political/outlooks.htm" (Accessed 3/7/2006).

[8] G. Sanders, A. Kadam “Prepubescent children show the adult relationship between dermatoglyphic asymmetry and performance on sexually dimorphic tasks.”, Cortex., 37(1):91-100, Feb, 2001.

[9] R. Austin, C. Reedy, “Implications of the IDENT/IAFIS Image Quality Study for Visa Fingerprint Processing”, Mitertek Systems (MTS), October 31, 2002

[10] G. Wendt , “Über weiße Linien im Abdruck der menschlichen Fingerbeere.” Homo 6:180–188, 1955.

[11] R. Gonzalez, R. Woods, Eddins: “Digital Image Processing.” Prentice Hall (2004)

[12] L. Hong, Y. Wang, and A. Jain. “Fingerprint image enhancement: Algorithm and performance evaluation.” Transactions on PAMI, 21(4):777–789, August 1998.

[13] A. Badawi, K. Hasan, E. Aly, and R. Messiha, "Chromosomes Classification Based on Neural Networks, Fuzzy Rule Based, and Template Matching Classifiers," In Proceedings of The 46th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), 2003.

[14] E. Lim, X. Jiang and W. Yau,"Fingerprint Quality and Validity Analysis",IEEE ICIP 2002.

[15] S. Chikkerur and V. Govindaraju,”Fingerprint Image Enhancement Using STFT Analysis”. International Workshop on Pattern Recognition for Crime Prevention, Security and Surveillance (ICAPR 05), pages 20–29, 2005.

[16] L. Hong, A.K. Jain, S. Pankanti, and R. Bolle, “Fingerprint Enhancement,”Proc. First IEEE WACV, pp. 202-207, Sarasota, Fla.,1996.

[17] M. Sonka, V. Halvac, and R. Boyle,” Image processing, Analysis and machine Vision”, Chapman & Hall, 1993*pan,* p. 301, 1982].